Math 113 2020-Sep-16

Topic: Boxplot

Example of boxplot. $n = 81$

want: min, 1st quartile, median, 3rd quartile, max

① sort data

② min = 9
max = 103

③ median M is data point numb $\frac{n+1}{2}$

$\frac{n+1}{2} = \frac{81+1}{2} = 41$ want #41

M = #41 = 70

④ The 1st quartile q_1 is the median of everything below the median.

(#1 thru #40) (#41)

find median. $\frac{40+1}{2} = 20.5$

q_1 = #20.5 = average of #20, #21

$= (52+54)/2 = 53$

③ Find q_3 the third quartile, which is median of ~~everything~~ above the original median

#41

#42 thru #81

renumber

"#1" thru "#40"

$\frac{40+1}{2} = 20.5$

q_3 = "#20.5" = average of "#20", "#21"

$= (92+93)/2 = 92.5$

+41

We renumbered #42 as our new "#1"

so

#61.5

"#20.5"

+41

20.5+41

Conclusion.
To find q_3... q_3 = # (from q_1) + one number below the rest of upper half

we just derived this rule.

In example above:

$\frac{40+1}{2} = 20.5$

$\frac{81+1}{2} = 41$

#1 = 9

q_1 = #20.5 = (52+54)/2 = 53

#40 =

M = #41 = 70

#42 =

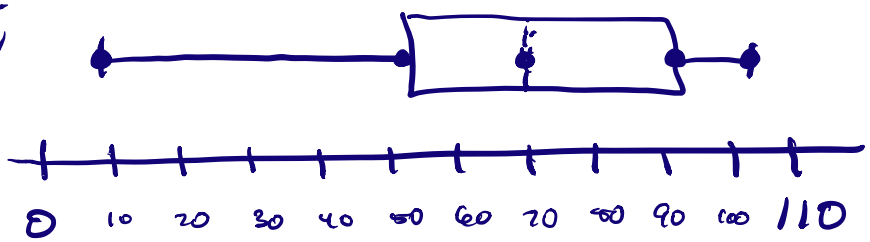
q_3 = #61.5 = (92+93)/2 = 92.5

#81 = 103

$42 - 1 + 20.5$

$= 61.5$

Box Plot



(like Figure 2.11 on page 97)

NOTE The text does not explain how to find q_1, q_3

Example 2.

Number of data pairs $n = 59$.

→ order/organize the data set

min = #1 = 6

q_1 = #15 = 6 $\left(\frac{29+1}{2}\right) = 15$

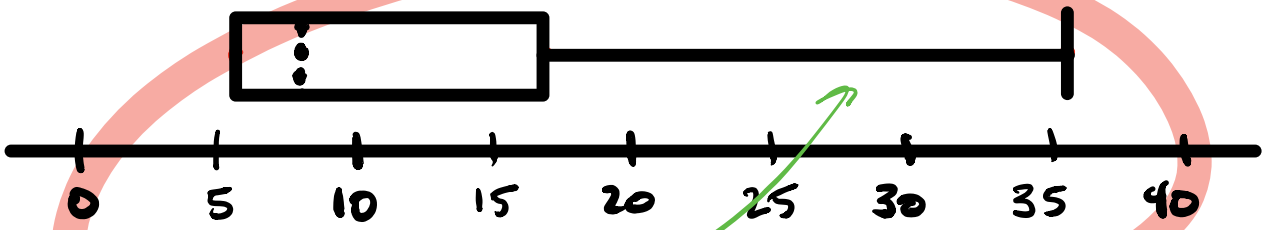
#29

M = #30 = 8 $\left(\frac{59+1}{2}\right) = 30$

#31 =

q_3 = #45 = 17 $\left(\frac{59+31}{2}\right) = 45$

max = #59 = 36



we call this a whisker
(usually there are 2)

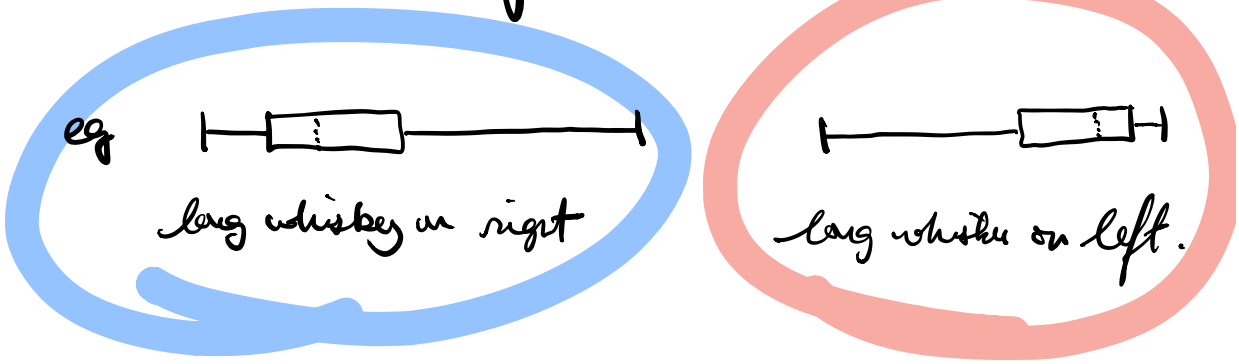
this data is skewed right
(that means we have outliers on the right)
ie, a few surprisingly large values

§ Skew

~~A long tail is~~

Data that makes a histogram with a long tail

will tend to make a box plot with a long whisker



we call these

right skewed (skewed right)

left skewed (skewed left)

Example 3

Given $n = 52$.

→ sort.

→

min = #1 = 70

1st quartile q_1 = #13.5 = 86

#26

Median M = #26.5 = 91

#27

3rd quartile q_3 = #39.5 = 94

max = #52 = 96

$\frac{1+26}{2} = 13.5$

q_1 = #13.5 = average of #13, #14

86, 86

$\frac{1+52}{2} = 26.5$

#26.5 = $\frac{\#26 + \#27}{2}$

$= \frac{91+91}{2} = 91$

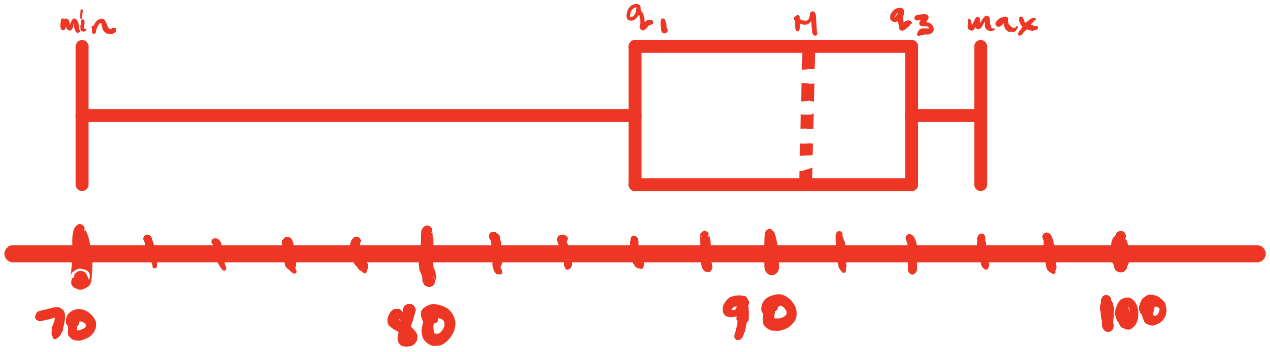
(the average of #26, #27)

#39.5 is the average of #39 = 94 and #40 = 94

$\frac{27+52}{2} = 39.5$

The 5 number summary

min, q_1, M, q_3, max = 70, 86, 91, 94, 96



This is left-skewed
(The data)
(The box-plot)